IN THE SPECIFICATION:

The specification is changed as follows:

Page 1, after the title and before line 1, insert

CROSS-REFERENCE TO RELATED APPLICATIONS

This Application is a Divisional of U.S. Application No. 09/263,692, filed March 5, 1999, the disclosure of which is incorporated herein by reference.

Please amend the paragraph bridging pages 7-8 as follows:

In an embodiment of the invention, a chemically synthesized promoter can comprise ~~of~~ a minimal domain (a) as depicted in SEQ ID NO. 2 (for high level expression of genes, i.e., strong promoter) or SEQ ID NO:3 (for low level expression of genes, i.e., weak promoter) and their derivatives comprising ~~of~~ variations as seen in Tables 1 and 2 respectively, functioning as TATA contexts in reference to artificial promoter falling between the positions -26 to -43 (The numbering of nucleotides is such that +1 indicates the first nucleotide of the transcription start site).

Please amend third paragraph on page 9 as follows:

In another embodiment of the invention, the chemically synthesized artificial promoter further comprises SEQ ID ~~NOs.~~NO:14 (for high level expression of genes, i.e., strong promoter) and SEQ ID NO:15 (for low level expression of genes, i.e., weak promoter) and their derivatives comprising ~~of~~ variations as seen in Tables 4 and 5 respectively, functioning as consensus sequences around the ATG start codon falling between the positions +83 to +102.

Please amend the second paragraph on page 12 as follows:

Computational analysis was carried out using the software from PC-Gene and database release 18-0 from Oxford Molecular Biology Group, Switzerland. A plant database comprising entries from plant genes only was created from the database CDEM 46 IN. It had 13,393 nucleic acid sequences. Depending on resemblance to a putative motif in the TATA and ATG regions, identified by comparing homology among 36 known highly expressed genes in plants, the database was classified into 262 transcriptionally highly expressed genes. Conserved motifs around the TATA region (Tables 1 and 2), transcriptional start site (Table 3) and translation initiation codon ATG (Tables 4 and 5) were identified for highly (Tables 1, 3 and 4) and lowly (Tables 2 and 5) expressed genes. The databases were then screened for possible conserved domains in the promoter region and further upstream of the coding region (reading frame) of genes. The highly conserved motif sequences along with the relatively less conserved regions and their variations to the extent seen in the Tables 1 and 5 gave characteristic component sequences that were assembled to develop an artificial promoter. The most highly conserved individual sequence motifs were identified as SEQ ID NO:2 to SEQ ID NO:16, and assembled to obtain the promoter regulatory sequence SEQ ID NO:1~~The individual motif sequences, most highly conserved were identified as ID SEQ 2 to ID SEQ 16 and assembled to obtain the promoter regulatory sequence ID SEQ 1~~.

Please amend the paragraph bridging pages 14-15 as follows:

A minimal promoter in eukaryotes is the DNA sequence proximal to the transcription initiation site. It usually contains an initiator *cis* element typically located ~30 nucleotides ~~nucleotide~~ upstream of the transcription start site (Aso, et al., J. Biol. Chem. 269: 26575-26583, 1994). The minimal promoter mainly consists of a sequence commonly called the ~~as~~ TATA element. Modulation of the formation or stability of the initiation complex by *trans-*

acting proteins that bind to distal *cis* elements ~~element~~ requires an intact TATA box (Horikoshi, et al., Cell 54: 665-669, 1998). Zhu, et al., (The plant cell 7: 1681-1689, 1988) showed ~~shown~~ TATATTTAA as a functional TATA box for the phenylalanine ammonia-lyase (PAL) promoter. *In vitro* studies conducted by Mukumoto, et al., (Plant Mol. Biol. 23: 995-1003) showed TATATATA as the sequence required for the plant TATA box. Till date, it is not known if TATATATA can be used as the minimal promoter in plants for expression of transgenes. Moreover, the minimal domain (a) used in this study and as depicted in SEQ ID No. 2 is different from those described in the earlier studies. All promoters in the database, as summarised in Table I have sequence motifs representing ~~ID~~ SEQ ID NO:2 or its variants within statistically insignificant limits. Table I represents the characteristic feature of TATA in highly expressed genes and the variation in the TATA region as noticed in different genes. The sequence domain as shown in SEQ ID No. 2 is (T/C)T(T/A)(T/C)NTCACTATATATAG ~~T3(T/A)TNTCACTATATAG (where T_3 indicates TTT appears at that site and~~ N indicates any one of the four nucleotides A,T,G or C can appear at that site) and is referred to as minimal domain (a) with respect to artificial synthetic promoter in this study. Our analysis of the database shows that the position of the sequence identified by us can vary from 40 to 28 nt upstream of the transcription start site. The lowly expressing genes show, the TATA consensus as $T_3N_4T_2$TATANNNAT ~~NT3N4T2TATANNNAT~~ (SEQ ID NO:~~No.~~3) which differs significantly from that found in consensus SEQ ID NO:~~No.~~2, and identified by us as a characteristic sequence in highly expressed genes. Thus the selection of sequence of TATA consensus region and its distance from the transcription start site may determine the level of gene expression. Mukumoto, et al., Plant Mol. Biol. 23: 995-1003 (1993) and Keith and Chua EMBO J.; 5 : 2419-2425 (1986) deduced the role of the TATA element by experimental evaluation. Their results

established the requirement of a sequence with certain critical nucleotide positions within the TATA element. Mutations at different positions were reported to reduce the activity of promoter considerably. An optimized TATA consensus sequence should be situated at a certain distance from the transcription initiation site for efficient initiation of transcription. A less than proper distance of the TATA element from transcription start site and a widely different variant TATA box sequence can be reduce expression as shown by Zhu, et al., The Plant cell, 7:1681-1689 (1995). Efficient recognition of the TATA element by TBP and TAF (TBP associating factors) regulatory factors determines the efficiency of transcription by RNA polymerase II. Our results identify a distinct sequence that can be employed to express genes in plants.

Please amend the paragraph encompassing lines 12-17 on page 16 as follows:

5' CCACTTGACG CACAATTGAC GCACAATGAC GCCACTTGAC GCTACT ~~CCACTTGACG CACAATTGAGCACAATACGCCACTTGACGCTACT~~ 3' (SEQ ID NO:~~No.~~5)

which may act as part of the minimal promoter, both in the sense as well as the antisense direction. Functional activity of the sequence constructed by us by employing a mix of C(C/A) (C/A) (A/C) T and TGACG either in prokaryotes or in eukaryotes and especially in plant cells is a novel part of this invention.

Please amend the last paragraph on page 17 as follows:

Domain I(a) somewhat resembles ~~the~~, but is different from the GC box reported by Menkens, et al., TIBS 20: 506-510 (1995) and may play ~~the~~ a role in the kinetics of opening of the transcription bubble and keeping the minimal promoter in a most active form to enhance transcription reinitiation from the transcription complex at the minimal promoter as suggested by Yean and Gralla, Nucl. Acids[.] Res. 24(14): 2723-2729 (1996). The functional

element~~domain I (a)~~ designed by us is duplicated and is different from any of the earlier reported sequences ~~sequence~~ and was predicted theoretically on the basis of computational analysis, as a possible efficient domain.

5' CACGTGCACGCGT 3' (SEQ ID NO:18)

The number of copies that could contribute to enhancing expression could vary, though three copies were taken by us as an example to demonstrate the principle.

Please amend the first paragraph on page 18 as follows:

Domain I (b) is also designed to be a trimer of the GATA type cis-acting element, as set forth in SEQ ID NO:19.

5' GATAGATAGATA 3' (SEQ ID NO:19)

The GATA elements are known to associate with the CaMV 35S promoter as shown by Odell, et al., Nature, 313: 810-812 (1985). On the basis of computational analysis, we predict this as a sequence that can be used in combination with other sequences to achieve a high level of transcription. The number of copies has been taken as three as an example, to demonstrate the principle and may be variable.

Please amend the second paragraph on page 18 as follows:

Domain I (c) is yet another artificial dimeric combination of the GTACGC type of element,~~elements,~~ as set forth in SEQ ID NO:20, noticed by us as commonly present in the region of -126 to -114 but less commonly present in the region of -90 to -120 nt.

5' GCTTGTACGCTGTACGCTGAC 3' (SEQ ID NO:20)

The GTACGC type of element has~~elements have~~ been described as the U box by Plesse, et al., ~~(1997)~~ Mol. Gen. Gent. 254: 258-266 (1997). We have included two such elements in the promoter designed in this study~~,~~ only as an example. The number of copies that contribute to improved function may be variable.

Please amend the first full paragraph on page 20 as follows:

Another 16 base pair palindromic sequence, 5' AC(G/A)(T/C)AAGCGCTTACGT ~~ACGTAAGCGCTTACGT~~ 3' (SEQ ID NO:~~No.~~10), is the octopine enhancer type of element and it's variants, which may or may not be palindromic. These were identified during this study to be conserved in several highly expressed plant genes and termed as domain II(d). This element was located more usually around -200 bp upstream. It may be active in both sense and antisense directions. The activity of the natural ocs element was shown by Gelvin, et al. Proc. Natl. Acad. Sci. USA, 85: 2553-557 (1988). However its use in association with other elements to develop a synthetic promoter is a novel aspect of this invention.

Please amend the first full paragraph on page 21 as follows:

The region between the transcription start site and the TATA box is also highly conserved and was identified by comparing several highly expressed genes. This region, viz.,

5' GGAAGTTCAT TTCATTTGGA ATGGACA
~~GGAGGTTCATTTCATTTGGATTGGACA~~ 3' (SEQ ID NO:~~No.~~12)

has not been identified earlier. It does not exactly resemble any known promoter and was computed purely by analysing the highly expressing genes and comparing the sequences with lowly expressed genes. Its length~~distance~~ varies between 20-40 nucleotides but usually is around 26 bp. This DNA sequence may function by lowering the Tm, and hence is predicted to facilitate transcription bubble formation and increase transcription efficiency. To that extent, the use of this element as well as its variants with lower Tm (AT richness) is a part of the new principle employed by us in developing an artificial promoter.

Please amend the paragraph bridging pages 22-23 as follows:

We also compared the translation initiation codon AUG context (that determines the ribosome halting at AUG and initiation complex formation) among highly and lowly

expressed genes. Improper context leads to bypassing of AUG by ribosomes, as shown by Kozak, J. Mol. Biol, 196: 947-950 (1987). We identified different contexts in different groups of plant genes which show significant differences in expression. The highly expressed genes show

AT(A/C)AACAATGGCTNCCNCNA (SEQ ID No. 14)

in contrast to the lowly expressed genes in plants which show

GANATGGNGNNGNNANA ~~GANATGNGNNGNNANA~~ (SEQ ID NO:~~No.~~15)

(Tables 4 & 5). SEQ ~~ID~~ ID NO:~~No.~~15 (although does not contain G after ATG). This indicated that the differences in the AUG context may be critical to achieve the desired level of gene expression. Analysis of the highly expressed genes, as seen in Table 4 suggests that the former sequence and its close variants allow high level expression of genes in nature. Hence, an artificial promoter targeted for high level of gene expression can have SEQ ID NO:~~No.~~14 or its variants to the extent given in Table 4.